

VU Research Portal

Approximate relativistic electron structure methods based on the quaternion modified Dirac

Visscher, L.; Saue, T.

published in

Journal of Chemical Physics
2000

DOI (link to publisher)

[10.1063/1.1288371](https://doi.org/10.1063/1.1288371)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Visscher, L., & Saue, T. (2000). Approximate relativistic electron structure methods based on the quaternion modified Dirac. *Journal of Chemical Physics*, 113, 3996-4002. <https://doi.org/10.1063/1.1288371>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Approximate relativistic electronic structure methods based on the quaternion modified Dirac equation

Lucas Visscher^{a)}

*Department of Theoretical Chemistry, Vrije Universiteit Amsterdam, Faculty of Sciences,
De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands*

Trond Saue

Department of Chemistry, University of Tromsø, N-9037 Tromsø Norway

(Received 13 April 2000; accepted 14 June 2000)

New implementations of the Lévy–Leblond, zeroth-order regular approach (ZORA) and spin-free Dirac equation are presented within the framework of the four-component relativistic program system DIRAC. This implementation allows systematic incorporation of relativistic effects at different levels of theory and corresponding computational cost. One of the possibilities of the new code is to neglect the effect of spin–orbit coupling in the orbital optimization process and introduce it in a later stage of the calculation. This method is shown to be unstable despite the boundedness of the spin–orbit operator itself. © 2000 American Institute of Physics. [S0021-9606(00)30434-2]

INTRODUCTION

It is well-known that relativity can have a profound impact on the electronic structure of molecules that contain heavy elements. Inclusion of these so-called relativistic effects in electronic structure calculations is usually done by means of an approximate relativistic Hamiltonian^{1–5} which consists of the nonrelativistic Hamiltonian plus additional operators to describe the dominant relativistic effects. Most of these Hamiltonians are written in such a way that one can exclude spin–orbit coupling operators and work with one-component instead of two-component wave functions. In the latter case one can then adapt efficient implementations of nonrelativistic electronic structure methods by making marginal changes in the one- and two-electron integral evaluation routines. The wave functions obtained in this fashion provide a basis for the introduction of spin–orbit coupling by means of degenerate perturbation theory or configuration interaction.

An alternative is to use the fully relativistic Dirac Hamiltonian that intrinsically describes both scalar relativistic effects and spin–orbit coupling. The advantage of this approach is that the incorporation of relativistic effects is done without approximations. Disadvantages are the higher computational cost of the solution methods and the additional implementation work that is required. In the last decade a number of implementations of the Hartree–Fock method have appeared but, in comparison to the nonrelativistic machinery available, implementations of electron correlation methods are still scarce. Part of this difference is caused by the necessity to use a double group symmetry formalism when utilizing molecular symmetry. Existing implementations of electron correlation methods cannot be used and new algorithms need to be developed. This complication is not due to the four-component character of the wave function but

is a consequence of the variational inclusion of spin–orbit coupling operators in the Hartree–Fock procedure. The same situation is encountered in two-component calculations and sometimes the machinery developed for four-component calculations has therefore been used in these cases as well.⁶

We will now, however, do the opposite and apply a non-variational, perturbative, treatment of spin–orbit coupling to four-component calculations. Our starting point is the modified four-component Hamiltonian introduced by Dyall.⁷ This Hamiltonian has the same eigenvalue spectrum as the conventional Dirac Hamiltonian but has solutions for which the upper and lower components share the same parity. Like the above-mentioned two-component approaches, Dyall's equation can be partitioned in a spin-free and a spin-dependent part. Since transformation to an approximate two-component relativistic scheme is avoided, this partitioning is, however, more rigorous than the usual approaches in which the scalar and spin–orbit operators are correct only to a certain order in an expansion parameter.

Besides reporting our study of valence-only spin–orbit approaches, this article also serves to introduce a new formalism for treating relativity in electronic structure methods. We combine Dyall's approach with the use of quaternion algebra⁸ and partitioning of the metric⁹ to develop a general formalism that incorporates nonrelativistic theory,¹⁰ the zeroth-order regular approach (ZORA), and the Dirac theory as special cases.

THE QUATERNION MODIFIED DIRAC EQUATION

A quaternion number q consists of one real and three imaginary parts. Using a vector notation for the quaternion imaginary part we write this as

$$\begin{aligned} \mathbf{q} &= {}^Rq + {}^I\mathbf{q}, \\ {}^Rq &= s, \\ {}^I\mathbf{q} &= \mathbf{v}_x\check{i} + \mathbf{v}_y\check{j} + \mathbf{v}_z\check{k}. \end{aligned} \quad (1)$$

^{a)}Author to whom correspondence should be addressed; electronic mail: Visscher@chem.vu.nl

Multiplication of quaternions obeys

$$\begin{aligned}\mathbf{q}_3 &= \mathbf{q}_1 \mathbf{q}_2, \\ s_3 &= s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2, \\ \mathbf{v}_3 &= s_1 \mathbf{v}_2 + \mathbf{v}_1 s_2 - \mathbf{v}_1 \times \mathbf{v}_2,\end{aligned}\quad (2)$$

and is noncommutative, as seen by the presence of the vector product in the quaternion imaginary part. These commutation rules bear close similarity to those of Pauli spin matrices and this relationship can be used to reformulate equations that have been written in terms of Pauli spin matrices. Saue and Jensen⁸ applied a quaternion transformation to derive a two-component quaternion Hamiltonian that is equivalent to the conventional four-component Dirac Hamiltonian. The corresponding Dirac equation (in Hartree atomic units) is

$$\hat{H}\Psi = E\Psi, \quad (3)$$

with

$$\hat{H} = \begin{pmatrix} {}^R\hat{V} & -c\,{}^I\hat{\mathbf{d}} \\ -c\,{}^I\hat{\mathbf{d}} & {}^R\hat{V} - 2c^2 \end{pmatrix}. \quad (4)$$

The superscripts R and I are used to identify operators that have only a real or quaternion imaginary part, respectively. The quaternion gradient operator

$${}^I\hat{\mathbf{d}} = \left(\check{i} \frac{\partial}{\partial z} + \check{j} \frac{\partial}{\partial y} + \check{k} \frac{\partial}{\partial x} \right) \quad (5)$$

plays the role that $(\sigma \cdot \hat{\mathbf{p}})$ has in the four-component formulation. We note in particular that the nonrelativistic kinetic energy operator is obtained as $\hat{T} = ({}^I\hat{\mathbf{d}} \cdot {}^I\hat{\mathbf{d}})/2$. The quaternion wave function is related to the four-component wave function $(\Psi^{L\alpha}, \Psi^{L\beta}, \Psi^{S\alpha}, \Psi^{S\beta})^T$ via

$$\begin{pmatrix} \Psi^L \\ \Psi^S \end{pmatrix} = \begin{pmatrix} \Psi^{L\alpha} - \Psi^{L\beta*} \check{j} \\ \Psi^{S\alpha} - \Psi^{S\beta*} \check{j} \end{pmatrix}. \quad (6)$$

The modified Dirac equation⁷ is obtained by defining a quaternion pseudolarge component ϕ^L that satisfies the equation

$$\Psi^S = \frac{-1}{2c} {}^I\hat{\mathbf{d}} \phi^L, \quad (7)$$

and reads

$$\begin{aligned} & \begin{pmatrix} {}^R\hat{V} & {}^R\hat{T} \\ {}^R\hat{T} & -{}^R\hat{T} + \frac{{}^I\hat{\mathbf{d}}^R V \cdot {}^I\hat{\mathbf{d}}}{4c^2} \end{pmatrix} \begin{pmatrix} \Psi^L \\ \phi^L \end{pmatrix} \\ &= E \begin{pmatrix} 1 & 0 \\ 0 & \frac{{}^R\hat{T}}{2c^2} \end{pmatrix} \begin{pmatrix} \Psi^L \\ \phi^L \end{pmatrix}. \end{aligned} \quad (8)$$

This equation can be split in real and quaternion imaginary parts when we define

$$\begin{aligned} {}^I\hat{\mathbf{d}}^R \hat{V} {}^I\hat{\mathbf{d}} &= {}^R\hat{W} + {}^I\hat{\mathbf{W}}, \\ {}^R\hat{W} &= -\hat{\nabla} \hat{V} \cdot \hat{\nabla}, \\ {}^I\hat{\mathbf{W}} &= -\check{i}(\hat{\nabla} \hat{V} \times \hat{\nabla})_z - \check{j}(\hat{\nabla} \hat{V} \times \hat{\nabla})_y - \check{k}(\hat{\nabla} \hat{V} \times \hat{\nabla})_x. \end{aligned} \quad (9)$$

Partitioning into real and quaternion imaginary parts corresponds to separating the spin-orbit coupling operators from the scalar relativistic operators. This is a very convenient feature of the quaternion formalism. The remaining real part of the equation can be partitioned in the spirit of direct perturbation theory⁹ in which the nonrelativistic Lévy-Leblond equation¹⁰ is used to define the zeroth-order problem. We choose to use three independent parameters for the Hamiltonian and metric and define

$$\mathbf{H}^{\text{MD}}(\lambda, \mu) \Psi^{\text{MD}} = E S^{\text{MD}}(\nu) \Psi^{\text{MD}}, \quad (10)$$

with

$$\mathbf{H}^{\text{MD}}(\lambda, \mu) = {}^R H^0 + \lambda {}^R H^1 + \mu {}^I \mathbf{H}^1, \quad (11)$$

and

$${}^R H^0 = \begin{pmatrix} \hat{V} & \hat{T} \\ \hat{T} & -\hat{T} \end{pmatrix}, \quad (12)$$

$${}^R H^1 = \frac{1}{4c^2} \begin{pmatrix} 0 & 0 \\ 0 & {}^R\hat{W} \end{pmatrix}, \quad (13)$$

$${}^I \mathbf{H}^1 = \frac{1}{4c^2} \begin{pmatrix} 0 & 0 \\ 0 & {}^I\hat{\mathbf{W}} \end{pmatrix}. \quad (14)$$

The metric is always real,

$${}^R S(\nu) = {}^R S^0 + \nu {}^R S^1, \quad (15)$$

with

$${}^R S_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (16)$$

and

$${}^R S^1 = \frac{1}{2c^2} \begin{pmatrix} 0 & 0 \\ 0 & \hat{T} \end{pmatrix}. \quad (17)$$

The parameters λ , μ , ν have the value one in the modified Dirac equation. Below we will discuss the approximate relativistic Hamiltonians that are found when choosing one or more parameters zero.

The wave function in Eq. (8) depends on the value of all three parameters λ , μ , and ν . The parameter μ is special because the modified Dirac equation contains only real operators if μ is taken zero. This means that the wave function can be chosen as a real function and, more importantly, that it transforms according to the irreducible representations of the molecular *single* point group. This makes the machinery developed for symmetry handling (both molecular point groups and spin-restricted formalisms) in nonrelativistic calculations applicable to the spinfree Dirac equation.

With three parameters that can assume the values one and zero we have eight possible combinations. Table I gives an overview of the combinations that are related to published

TABLE I. Special cases for different values of the parameter set.

λ	μ	ν	Equation
1	1	1	Modified Dirac ^a
1	0	1	Spinfree Modified Dirac ^a
1	1	0	ZORA ^b
1	0	0	Spinfree ZORA ^b
0	0	0	Lévy–Leblond ^c

^aReference 7.^bReference 5.^cReference 10.

relativistic equations. The compact quaternion formulation outlined above makes it easy to combine these different relativistic equations in one computer program. Such an implementation allows the use of more approximate, cheaper, schemes in the initial stages of an iterative or step-wise calculation which can be used to increase its efficiency. The formalism also provides a convenient starting point for the formulation of perturbation theory. We will now discuss the special features of the equations listed in Table I and describe their implementation in the program system DIRAC.¹¹

IMPLEMENTATION OF THE QUATERNION MODIFIED DIRAC EQUATION IN THE DIRAC PROGRAM SYSTEM

Expansion of the quaternion Dirac equation in a finite (generally nonorthogonal) basis of real four-component functions $\{\chi^L, \chi^S\}$ leads to a eigenvalue problem of the form

$$\mathbf{F}\mathbf{c} = \epsilon\mathbf{S}\mathbf{c}, \quad (18)$$

which is solved by first transforming to an orthonormal basis $\{\theta^L, \theta^S\}$ via a Löwdin transformation

$$\begin{aligned} \mathbf{F}'\mathbf{c}' &= \epsilon\mathbf{c}', \\ \mathbf{F}' &= \mathbf{A}^{-1}\mathbf{F}\mathbf{A}, \end{aligned} \quad (19)$$

$$\mathbf{c}' = \mathbf{A}^{-1}\mathbf{c},$$

with the transformation matrix \mathbf{A} defined by

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{S}^{-1/2}, \\ \mathbf{U}^\dagger\mathbf{S}\mathbf{U} &= \mathbf{s}, \\ s_{ij} &= s_i\delta_{ij}. \end{aligned} \quad (20)$$

Since the overlap matrix is real and blocked on the large and small components the transformation matrix will also be real and blocked. Explicitly we have

$$\begin{aligned} |\theta_\mu^L\rangle &= \sum_\nu^{n_L} |\chi_\nu^L\rangle^R \mathbf{A}_{\nu\mu}^L, \\ |\theta_\mu^S\rangle &= \sum_\nu^{n_S} |\chi_\nu^S\rangle^R \mathbf{A}_{\nu\mu}^S. \end{aligned} \quad (21)$$

The transformation from the Dirac equation to the modified Dirac equation involves a change of metric and is described by a nonunitary quaternion transformation matrix. This matrix may be combined with the Löwdin transformation ma-

trix to give one compound transformation matrix \mathbf{B} . We do this by defining a new set of small component basis functions $\{\vartheta^S\}$ as

$$\begin{aligned} |\vartheta_\mu^S\rangle &= \frac{I\mathbf{d}}{-2c} |\theta_\mu^L\rangle = \sum_\alpha^{n_S} |\theta_\alpha^S\rangle \left\langle \theta_\alpha^S \left| \frac{I\mathbf{d}}{-2c} \right| \theta_\mu^L \right\rangle \\ &= \sum_\nu^{n_S} |\chi_\nu^S\rangle^I \mathbf{B}_{\nu\mu}^S, \\ \mathbf{B}_{\nu\mu}^S &= \sum_\alpha^{n_S} {}^R\mathbf{A}_{\nu\alpha}^S \left\langle \theta_\alpha^S \left| \frac{I\mathbf{d}}{-2c} \right| \theta_\mu^L \right\rangle, \\ \mathbf{B} &= \begin{pmatrix} {}^R\mathbf{A}^L & 0 \\ 0 & {}^I\mathbf{B}^S \end{pmatrix}, \end{aligned} \quad (22)$$

and obtain the small-component part of \mathbf{B} . The large component part is identical to the large component part of the original Löwdin matrix \mathbf{A} . We note that the resolution of identity that is used is only complete if $\{\theta^S\}$ spans all functions that are generated by applying the quaternion gradient operator on the set $\{\theta^L\}$. This is the kinetic balance condition that is fulfilled because we choose the primitive small component basis set as $\{\chi^S\} = \{(\partial/\partial x)\chi^L\} \cup \{(\partial/\partial y)\chi^L\} \cup \{(\partial/\partial z)\chi^L\}$. We can then simply replace the transformation matrix \mathbf{A} by \mathbf{B} and obtain a matrix representation of the modified Dirac equation. The kinetic energy matrix in the orthonormal large component basis $\{\theta^L\}$ is now identical to the overlap matrix in the nonorthogonal small component basis $\{\vartheta^S\}$ (multiplied by a factor $2c^2$)

$$\begin{aligned} \langle \vartheta_\mu^S | \vartheta_\nu^S \rangle &= \frac{1}{4c^2} \sum_{\alpha,\beta}^{n_S} \langle \theta_\mu^L | I\mathbf{d} | \theta_\alpha^S \rangle \langle \theta_\alpha^S | \theta_\beta^S \rangle \langle \theta_\beta^S | I\mathbf{d} | \theta_\nu^L \rangle \\ &= \frac{1}{2c^2} \langle \theta_\mu^L | T | \theta_\nu^L \rangle, \end{aligned} \quad (23)$$

and the potential energy matrix becomes

$$\begin{aligned} \langle \vartheta_\mu^S | V | \vartheta_\nu^S \rangle &= \frac{1}{4c^2} \sum_{\alpha,\beta}^{n_S} \langle \theta_\mu^L | I\mathbf{d} | \theta_\alpha^S \rangle \langle \theta_\alpha^S | V | \theta_\beta^S \rangle \langle \theta_\beta^S | I\mathbf{d} | \theta_\nu^L \rangle \\ &= \frac{1}{4c^2} \langle \theta_\mu^L | {}^R\mathbf{W} | \theta_\nu^L \rangle + \frac{1}{4c^2} \langle \theta_\mu^L | {}^I\mathbf{W} | \theta_\nu^L \rangle. \end{aligned} \quad (24)$$

The matrix representation of the modified Dirac equation is partitioned in the same way as described above. The final working equations are obtained after inclusion of the Coulomb interaction between the electrons in the Hartree–Fock approximation and read

$$\mathbf{F}(\lambda, \mu)\mathbf{c} = \epsilon\mathbf{S}(\nu)\mathbf{c}, \quad (25)$$

with

$$\mathbf{F}(\lambda, \mu) = \mathbf{H}(\lambda, \mu) + \mathbf{J}(\lambda, \mu) - \mathbf{K}(\lambda, \mu). \quad (26)$$

We hereby defined the Coulomb matrix \mathbf{J} as

$$\mathbf{J}_{\kappa\lambda}^{XY} = \sum_{\mu,\nu} 2\mathbf{D}_{\nu\mu}^{WV} \cdot \mathbf{I}_{\mu\nu,\kappa\lambda}^{VW,XY}, \quad (27)$$

and the exchange matrix \mathbf{K} as

$$\mathbf{K}_{\kappa\lambda}^{XY} = \sum_{\mu,\nu} \mathbf{D}_{\nu\mu}^{WV} \cdot \mathbf{I}_{\mu\lambda,\kappa\nu}^{VY,XW}, \quad (28)$$

with

$$\mathbf{I}_{\mu\nu,\kappa\lambda}^{VW,XY} = \iint \frac{\partial_{\mu}^{V\dagger}(r_1) \partial_{\nu}^W(r_1) \partial_{\kappa}^{X\dagger}(r_2) \partial_{\lambda}^Y(r_2)}{r_{12}} \times \delta_{VW} \delta_{XY} d\mathbf{r}_1 d\mathbf{r}_2, \quad (29)$$

and

$$\mathbf{D}_{\nu\mu}^{WV} = \sum_i \mathbf{C}_{\mu i}^{V*} \mathbf{C}_{\nu i}^W. \quad (30)$$

Choosing $\lambda=0$ implies that only the LL parts of the Coulomb and exchange matrices are kept. Choosing $\mu=0$ means that only the real parts of these matrices survive and that the spin-dependent terms, that is contributions to spin-orbit coupling, are eliminated.

The Dirac equation

The Dirac equation corresponds to the choice of parameters ($\lambda=1$, $\mu=1$, $\nu=1$). The integral-direct algorithm for solving the Hartree-Fock equations described in detail in Ref. 12 remains almost unmodified. The only difference is that the Löwdin transformation to the orthonormal basis is found via the intermediate basis defined in Eq. (22). This makes the contributions of spin-orbit terms appear explicitly, either as off-diagonal blocks coupling the distinct symmetry adapted diagonal blocks, or as (quaternion-)imaginary parts in an otherwise real matrix.

The spinfree Dirac method

Elimination of spin-orbit coupling terms while keeping the scalar relativistic terms is done by choosing the parameter set ($\lambda=1$, $\mu=0$, $\nu=1$). This is implemented in DIRAC by a call to a routine that zeroes the quaternion-imaginary blocks of the Fock matrix prior to diagonalization of this matrix. This is the only necessary change relative to an ordinary Dirac-Hartree-Fock calculation. The coefficient and density matrices automatically become real-valued because they are obtained via a diagonalization of a real matrix. In order to achieve additional computational savings we, however, also built a structure in which the blocks that belong to different boson irreps are diagonalized separately. This structure is not present in a normal Dirac-Hartree-Fock implementation in which the matrix is blocked according to double group symmetry considerations only. We can now identify the four-component spinors as belonging to a specific boson irrep and pass this information to post-Hartree-Fock programs that work in the molecular orbital basis. In these programs we can then use exactly the same treatment of molecular point group symmetry as employed in nonrelativistic implementations. It also means that any nonrelativistic electron correlation implementation that is formulated entirely on molecular orbital (MO)-basis can be interfaced and work with the spinfree Dirac equation.

The Lévy-Leblond method

The Lévy-Leblond method¹⁰ is obtained by choosing the parameter set ($\lambda=0$, $\mu=0$, $\nu=0$). The method gives four-component positive energy solutions but the negative energy solutions become undefined. This makes it necessary to apply a preprojection on the positive energy states. We do this by defining a two-component basis set that has a fixed ratio between the large and small components:

$$|\vartheta_{\mu}^{LL}\rangle = \begin{pmatrix} |\vartheta_{\mu}^L\rangle \\ |\vartheta_{\mu}^S\rangle \end{pmatrix} = \begin{pmatrix} |\vartheta_{\mu}^L\rangle \\ \frac{I\mathbf{d}}{-2c} |\vartheta_{\mu}^L\rangle \end{pmatrix}. \quad (31)$$

This gives a reduced transformation matrix \mathbf{B} ,

$$\varrho \mathbf{B}_{\nu\mu} = \begin{pmatrix} \mathbf{A}_{\nu\mu}^L & \sum_{\alpha}^{n_S} {}^R \mathbf{A}_{\nu\alpha}^S \left\langle \vartheta_{\alpha}^S \left| \frac{I\mathbf{d}}{-2c} \right| \vartheta_{\mu}^L \right\rangle \end{pmatrix}. \quad (32)$$

This basis set has now only half the dimension of the set defined in Eq. (22). It is easy to show that the resulting Hartree-Fock matrix equation is identical to the matrix representation of the nonrelativistic Hartree-Fock equation. The small component functions only serve as an auxiliary basis to express the kinetic energy operator as an inner product of momentum operator matrices.

The Dirac-Hartree-Fock implementation again required only very little modification to incorporate the Lévy-Leblond-type formalism. Besides transforming to the orthonormal basis via Eq. (32) we make the small component part of the nuclear attraction matrix zero and skip the calculation of the $(SS|LL)$ and $(SS|SS)$ classes of two-electron repulsion integrals. The implementations of post-Hartree-Fock methods did not need any modification.

The ZORA method

The zeroth-order regular approach (ZORA) is obtained by choosing the parameter set ($\lambda=1$, μ , $\nu=0$). This is clear by writing Eq. (8) in two-component form via elimination of the lower component

$${}^R \hat{V} \psi^L + \hat{\mathbf{T}}^{\text{ZORA}}(\mu) \psi^L = \psi^L E, \quad (33)$$

$$\hat{\mathbf{T}}^{\text{ZORA}}(\mu) = {}^R \hat{\mathbf{T}} ({}^R \hat{\mathbf{T}} - {}^R \hat{W} - \mu I \hat{\mathbf{W}})^{-1} {}^R \hat{\mathbf{T}}. \quad (34)$$

For $\mu=1$ we can recombine the terms in the inverse operator

$$\hat{\mathbf{T}}^{\text{ZORA}}(1) = 2 {}^R \hat{\mathbf{T}} \left(I \hat{\mathbf{d}} \left(1 - \frac{{}^R \hat{V}}{2c^2} \right) \cdot I \hat{\mathbf{d}} \right)^{-1} {}^R \hat{\mathbf{T}}, \quad (35)$$

to remove the momentum operator

$$\hat{\mathbf{T}}^{\text{ZORA}}(1) = \frac{1}{2} I \hat{\mathbf{d}} \left(1 - \frac{{}^R \hat{V}}{2c^2} \right)^{-1} I \hat{\mathbf{d}}. \quad (36)$$

This allows a new distinction between scalar and spin-dependent terms

$$\begin{aligned} \hat{\mathbf{T}}^{\text{ZORA}}(1) &= {}^R \hat{\mathbf{T}}^{\text{ZORA}} + I \hat{\mathbf{T}}^{\text{ZORA}}, \\ {}^R \hat{\mathbf{T}}^{\text{ZORA}} &= -\frac{1}{2} \hat{\mathbf{V}} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \cdot \hat{\mathbf{V}}, \end{aligned}$$

$${}^I T^{\text{ZORA}} = -\frac{1}{2} \check{i} \left(\hat{\nabla} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \times \hat{\nabla} \right)_z - \frac{1}{2} \check{j} \left(\hat{\nabla} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \times \hat{\nabla} \right)_y - \frac{1}{2} \check{k} \left(\hat{\nabla} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \times \hat{\nabla} \right)_x, \quad (37)$$

giving finally

$$\hat{T}^{\text{ZORA}}(1) = -\frac{1}{2} \hat{\nabla} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \cdot \hat{\nabla} - \frac{1}{4c^2} \left(1 - \frac{\hat{V}}{2c^2} \right)^{-2} (\check{i}(\hat{\nabla} \hat{V} \times \hat{\nabla})_z + \check{j}(\hat{\nabla} \hat{V} \times \hat{\nabla})_y + \check{k}(\hat{\nabla} \hat{V} \times \hat{\nabla})_x), \quad (38)$$

a standard formulation of the ZORA method cast in quaternion notation. One may use this formulation to make Eq. (33) spinfree by leaving out the quaternion imaginary part of Eq. (38). Note that this spinfree equation does not correspond to the limit $\mu=0$ because the transition from Eq. (34) to Eq. (35) is only possible for $\mu=1$. This is a new example of the ambiguity in defining scalar and spin-orbit effects.¹³

We have the same problem with zero eigenvalues of the metric matrix as encountered previously with the Lévy–Leblond method. An analogous solution is obtained by defining the reduced basis

$$|\vartheta_\mu^{\text{ZORA}}\rangle = \begin{pmatrix} |\vartheta_\mu^L\rangle \\ |\vartheta_\mu^S\rangle \end{pmatrix} = \begin{pmatrix} |\vartheta_\mu^L\rangle \\ \left(1 - \frac{\hat{V}}{2c^2} \right)^{-1} \frac{\hat{d}}{2c} |\vartheta_\mu^L\rangle \end{pmatrix}. \quad (39)$$

The difference between the two approaches is that the ZORA small component parts cannot be expressed exactly in terms of primitive Cartesian Gaussian-type orbitals (GTOs). When deriving the ZORA equation from the Dirac–Hartree–Fock operator,¹⁴ one furthermore has the complication that the Coulomb and exchange matrices appear in the relation between the large and small components of the wave function. Faas *et al.*¹⁴ therefore proposed to neglect picture change effects and use only the contribution to the potential that arises from the large component density. The only difference with the one-electron formalism is then a diagonal Coulomb contribution in the lower block of V that represents the shielding of the nuclei due to the electronic charge in the large components. This makes the use of Eq. (39) feasible but still laborious because this shielding contribution varies during the self-consistent field (SCF) process. This requires in principle reconstruction of all matrix elements in every iteration due to the induced change of basis. In addition they used a resolution of identity to evaluate integrals over one- or two-component operators.

We use a slightly different method that is based on the so-called full density ZORA-4 formalism. This is done using the same basis as chosen in the four-component calculations, i.e., defined via Eq. (22) instead of Eq. (39). Prior to diagonalization we invert the SS part of the matrix and use this matrix to obtain the Fock matrix appropriate for the ZORA

TABLE II. Orbital energies (a.u.) and fine-structure splitting (a.u.) for hydrogen-like uranium.

	$2p$	$2p^{1/2}$	$2p^{3/2}$	Splitting
Lévy–Leblond	−1058.0000			
SF–ZORA	−1172.3974			
SF–Dirac	−1136.9078			
ZORA		−1300.9223	−1122.1673	178.76
Dirac		−1257.3709	−1089.6114	167.76
Dirac (exact)		−1257.3959	−1089.6114	167.78

scheme. This procedure is identical to insertion of the resolution of identity in the lower component of Eq. (39). The iterative recalculation of matrix elements is thus in our scheme replaced by one additional diagonalization in each iteration that gives us the transformation from the fixed four-component basis towards the iterative two-component basis. On convergence both methods should give the same result. The approximate density scheme suggested by Faas *et al.*¹⁴ can also be obtained by putting the LS , SL , and SS blocks of the density matrix to zero. At present we implemented both the full and the approximate density scheme for closed-shell Hartree–Fock calculations. Incorporation in Kramers-restricted open-shell calculations is possible but more involved because the Fock matrix is not uniquely defined in restricted open-shell calculations.

Other schemes

Three binary parameters give eight possibilities for defining a new formalism. We described five different choices and identified them as methods that have been developed before. The three remaining options are of less practical interest. They concern methods in which scalar relativistic effects are neglected and only spin-orbit coupling is taken into account. This is not very efficient because it is the latter effect that is most difficult and costly to evaluate.

APPLICATIONS

The validity of the implementation was tested by a calculation on hydrogen-like uranium. We use the basis set given by Dyall⁷ and list the $2p$ orbital energies calculated by the five different methods in Table II. The results published by Dyall⁷ and van Leeuwen *et al.*¹⁵ are reproduced accurately. This gives us confidence that the approximate resolution of identity applied in the ZORA implementation does not give rise to significant errors if the primitive basis set is of reasonable size.

A more interesting application of the quaternion modified Dirac equation is the use of perturbation theory in the calculation of spin-orbit interactions. A popular method is the so-called SO–CI method¹⁶ where a configuration interaction calculation that includes spin-orbit matrix elements is based on orbitals obtained in a scalar relativistic calculation. This scheme can be used in any method that allows identification of a separate spin-orbit operator. Within the SO–CI approaches there is a distinction between “perturbational” methods in which the spin-orbit matrix elements are included after the CI-wave functions are determined and

TABLE III. Fine-structure splitting (cm^{-1}) of the 2P ground state of the group XIII elements. PT: Degenerate perturbation theory, SOCI-X: SO–CI calculation allowing single excitations from the valence p orbitals to virtual orbitals below X a.u. SOCI-All: SO–CI calculation allowing single excitations from the valence p orbitals to all virtual orbitals. Full: Hartree–Fock step done using the full Dirac Hamiltonian. Expt.: Experimental value of the splitting.^a

	PT	SOCI-10	SOCI-100	SOCI-All	Full	Expt.
B	20.34	20.35	20.35	20.35	20.35	15.25
Al	99.73	99.88	99.99	100.17	117.63	112.1
Ga	726.1	736.9	742.3	754.3	792.5	826.2
In	1944	2016	2082	2212	2156	2213
Tl	6499	7412	8613	14741	7643	7793

^aReferences 26, 27

“variational” methods in which these matrix elements are included in the CI procedure itself. A characteristic of the variational methods is that spin–orbit and electron correlation effects are included simultaneously but some higher-order mixing may also occur in the perturbational approaches.

We want to focus on possible artefacts caused by neglecting spin–orbit coupling effects in the orbital generation step. This is done by restricting the CI space to singly excited determinants. This excludes the differential electron correlation effects that could otherwise complicate the analysis. Within this restriction we can still simulate the essential features of both the perturbational and the variational SO–CI variants.

We take the fine structure splitting (FSS) of the 2P ground state of the thallium atom as a representative example. The 2P state is influenced considerably by both scalar relativistic and spin–orbit effects and has been a testing ground for many relativistic methods. We refer to the recent review by Hess and Marian¹⁷ for a good overview.

Since we focus on spin–orbit effects that find their origin in the region close to the nucleus it is of importance to include sufficiently tight functions. We use a large relativistically optimized $24s22p16d10f2g$ GTO-type basis set developed by Faegri¹⁸ in which the tightest p -function has exponent 2.10^7 . This is sufficient for the present purpose. In order to provide a flexible description of the core region we use this basis set in uncontracted form. To better understand the calculated values for Tl we also calculate the FSS of the lighter group XIII elements using the cc-pV5Z basis sets of Dunning and co-workers^{19–21} (elements B, Al, and Ga) and the relativistically optimized cc-pVDZ basis set of Dyal²² (In), also in uncontracted form. In all calculations we neglected the contribution of the $(SS|SS)$ -type of Coulomb two-electron integrals²³.

Table III shows that the SO–CI approach works quite well up to gallium. In this case it still underestimates the full splitting even when all virtual orbitals are taken into account. The SO–CI approach starts to overestimate the calculated splitting in indium if virtual orbitals with energies higher than 100 a.u. are taken into account and this overestimation becomes dramatic for thallium where the SO–CI-all value is almost twice the reference value.

The observed overestimation is caused by a pseudova-

TABLE IV. Total electronic energies (in a.u.) and fine-structure splitting (cm^{-1}) of the 2P ground state of thallium. Active orbitals: orbitals from which single excitations are allowed. The last line gives the result of the reference calculation in which SO-coupling is included in the Hartree–Fock step. $(SS|SS)$ type two-electron integrals were neglected.

Active orbitals	$E(^2P_{1/2})$	$E(^2P_{3/2})$	FSS
$6p$	–20 254.184 65	–20 254.117 48	14 741
$5p6p$	–20 255.170 85	–20 255.123 93	10 298
$4p5p6p$	–20 258.855 72	–20 258.815 39	8851
$3p4p5p6p$	–20 268.962 11	–20 268.925 53	8028
$2p3p4p5p6p$	–20 294.374 97	–20 294.341 01	7454
Full calculation	–20 277.514 01	–20 277.479 18	7643

riational collapse in the valence CI approach. The singly excited CI procedure gives an approximate relaxation of the active valence orbitals under the influence of the perturbing operator. This relaxation induces the largest changes in the core region where the spin–orbit operator has its maximum amplitude. The problem is now that the relaxed valence orbitals are kept orthogonal on the original core orbitals while they should become orthogonal on the true Dirac core orbitals. We call this a pseudovariational collapse problem because the relaxation can be interpreted as causing the valence orbital energies to collapse to energies far below the exact value. It is not a true variational collapse because the energies are still bound and because the total energy is still above the exact total energy.

This artifact can be remedied by including single excitations from the core p -orbitals into the valence p -orbitals. We illustrate this by a series of calculations in which we extend the active space with one extra shell of core orbitals at a time. From Table IV it is clear that even excitations from $2p$ orbitals give a nonnegligible effect on the calculated splitting. The effect on the total energy is less relevant for chemical applications but its magnitude shows the underlying cause of the sensitivity to core excitations.

While the origin of pseudovariational collapse is clear its solution is not obvious. Routine inclusion of excitations from core orbitals will not be feasible in most SO–CI calculations. An alternative is the use of cutoff threshold for the high virtuals that are causing this collapse. This is the solution that is used in most applications of the SO–CI method, either implicitly by not including tight functions in a valence only basis set, or explicitly via some kind of energy selection mechanism in the CI procedure. The present results show that this procedure can give only limited accuracy as one can—depending on the chosen threshold—either under- or overestimate the true splitting significantly.

The pseudovariational collapse problem that we showed for the spinfree Dirac equation should also be present also in other relativistic approaches like the Douglas–Kroll–Hess (DKH) method and the ZORA method if the basis set gives a flexible description of the core region. Rakowitz and Marian²⁴ and Wahlgren *et al.*²⁵ report effects in the order of a few hundred wave numbers in the FSS when varying the active space in all-electron DKH-type calculations but they both used generally contracted basis sets with contraction coefficients based on a atomic scalar DKH-type calculation.

This gives an essentially single zeta description of the deep core orbitals and eliminates the presence of tight orbitals in the virtual space. We anticipate that more dramatic effects will show up in DKH calculations with large uncontracted basis sets.

Most of the discussions about the SO–CI method applied to the thallium atom have focused on the large difference in radial expectation value $\langle r \rangle$ between the $6p_{1/2}$ and $6p_{3/2}$ orbitals. This difference is indeed considerable and important in the description of chemical bonding but it is not the most crucial property to consider when discussing the FSS. What matters is the difference in shape in the core region, which is heavily weighted by the spin–orbit operator. The properly relaxed $6p$ orbitals have core wiggles that secure orthogonality on the true $np_{1/2}$ and $np_{3/2}$ ($n < 6$) orbitals while the unrelaxed orbitals are orthogonal to the scalar relativistic $6p$ orbitals. The transition from the scalar relativistic $6p$ orbitals towards the $6p_{1/2}$ and $6p_{3/2}$ orbitals requires both the presence of tight functions in the virtual space as well as alteration of the core orthogonality condition. Imbalances in the description of these two requirements can lead to meaningless results.

CONCLUSIONS

We presented the implementation of the quaternion modified Dirac equation. This implementation offers possibilities for applying different relativistic electronic structure methods within the same theoretical and computational framework. This is one step on the road to tuneable calculations, i.e., approaches where the level of (relativistic) theory and computational cost is tuned to the desired accuracy.

We used this implementation to perform SO–CI calculations and showed that a valence-only approach can lead to severe overestimation of fine-structure splittings due to a pseudovariational collapse problem.

ACKNOWLEDGMENTS

L.V. thanks Dr. Erik van Lenthe and Dr. Ken Dyall for useful discussions.

- ¹L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
- ²M. Douglas and N. M. Kroll, Ann. Phys. **82**, 89 (1974).
- ³B. A. Hess, Phys. Rev. A **33**, 3742 (1986).
- ⁴C. Chang, M. Pélissier, and P. Durand, Phys. Scr. **34**, 394 (1986).
- ⁵E. van Lenthe, E. J. Baerends, and J. G. Snijders, J. Chem. Phys. **99**, 4597 (1993).
- ⁶H. S. Lee, Y. K. Han, M. C. Kim, C. Bae, and Y. S. Lee, Chem. Phys. Lett. **293**, 97 (1998).
- ⁷K. G. Dyall, J. Chem. Phys. **100**, 2118 (1994).
- ⁸T. Saue and H. J. Aa. Jensen, J. Chem. Phys. **111**, 6211 (1999).
- ⁹W. Kutzelnigg, Z. Phys. D: At., Mol. Clusters **11**, 15 (1989).
- ¹⁰J. M. Lévy-Leblond, Commun. Math. Phys. **6**, 286 (1967).
- ¹¹DIRAC, a relativistic *ab initio* electronic structure program, Release 3.1, written by T. Saue, T. Enevoldsen, T. Helgaker, H. J. Aa. Jensen, J. K. Laerdahl, K. Ruud, J. Thyssen, and L. Visscher (<http://dirac.chem.sdu.dk>), 1998.
- ¹²T. Saue, K. Fægri, T. Helgaker, and O. Gropen, Mol. Phys. **91**, 937 (1997).
- ¹³L. Visscher and E. van Lenthe, Chem. Phys. Lett. **306**, 357 (1999).
- ¹⁴S. Faas, J. G. Snijders, J. H. Van Lenthe, E. Van Lenthe, and E. J. Baerends, Chem. Phys. Lett. **246**, 632 (1995).
- ¹⁵R. van Leeuwen, E. van Lenthe, E. J. Baerends, and J. G. Snijders, J. Chem. Phys. **101**, 1272 (1994).
- ¹⁶B. A. Hess, C. M. Marian, and S. D. Peyerimhoff, in *Modern Electronic Structure Theory*, edited by D. R. Yarkoni (World Scientific, Singapore, 1995), Vol. 1, pp. 152.
- ¹⁷B. A. Hess and C. M. Marian, in *Computational Molecular Spectroscopy*, edited by P. Jensen and P. R. Bunker (Wiley, Sussex, 1999).
- ¹⁸K. Fægri, personal communication (1999).
- ¹⁹T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
- ²⁰D. Woon and T. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
- ²¹A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **110**, 7667 (1999).
- ²²K. G. Dyall, Theor. Chem. Acc. **99**, 366 (1998).
- ²³L. Visscher, Theor. Chem. Acc. **98**, 68 (1997).
- ²⁴F. Rakowitz and C. M. Marian, Chem. Phys. Lett. **257**, 105 (1996).
- ²⁵U. Wahlgren, M. Sjøvoll, H. Fagerli, O. Gropen, and B. Schimmelpfennig, Theor. Chem. Acc. **97**, 324 (1997).
- ²⁶Atomic Spectra Database Version 2.0 (1999), National Institute for Standards and Technology (<http://physics.nist.gov/asd>).
- ²⁷C. E. Moore, *Atomic Energy Levels Vol. III* (National Bureau of Standards Circular 467, Washington, 1958).